# European Soccer Transfer Market

Capstone #1 Slide Deck

Introduction

- Transfer fees paid and received by European soccer clubs have skyrocketed in the last decade

- There have been a number of clubs that have started to compete at the highest level after emerging from relative obscurity
  - Manchester City in England
  - Paris Saint-Germain in France

- Fees and revenues began to rise sharply in 2012

- An influx of outside money to certain clubs not only increased spending for these suddenly richer clubs, but also forced rivals to start spending more along with them
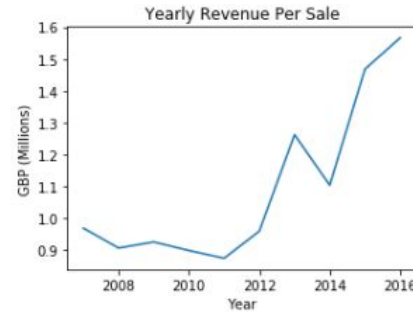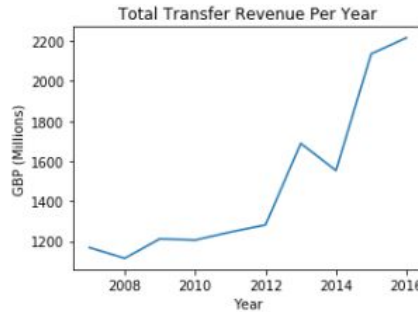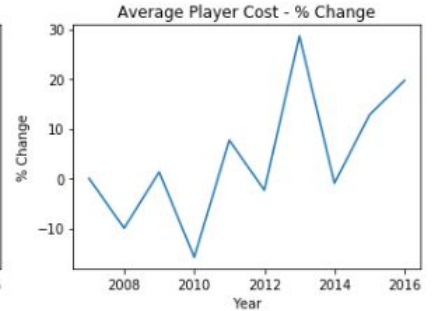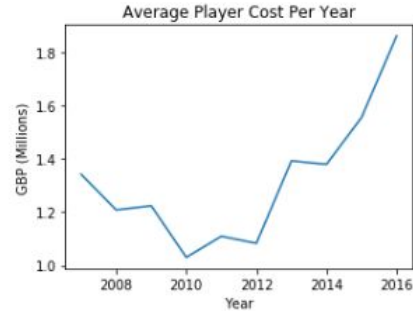
Problem

Does this massive increase in spending my European soccer teams actually translate to success on the field, or are there other factors that are more important to a team's success?

Data

- Match Data
  - https://www.kaggle.com/hugomathien/soccer
  - Contains every match result played by European teams between 08/09 and 15/16 seasons

- Transfer Data
  - https://github.com/ewenme/transfers/tree/master/data
  - Contains every dollar spent and received by European teams during the same time frame

- Target Data - Teams who did not get relegated to lower division between 08/09 and 15/16
  - More consistent budget due to TV deals, merchandise, wealthier owners

Data

- Joined each data set by team, so all of the match and transfer data for each team was contained in a single table

- Created the variable position value to equate success across all leagues
    - Position Value = 100 * 5(league finish - 1)
    - Starts at 100 for a first place finish, decreases by 5 with each lower position

- Goal is to see if position value increases with money spent

Relationship Between Average Position Value and Total Spent

- For each of the top seven leagues, there is a generally positive correlation between average position value and total spent during the time frame
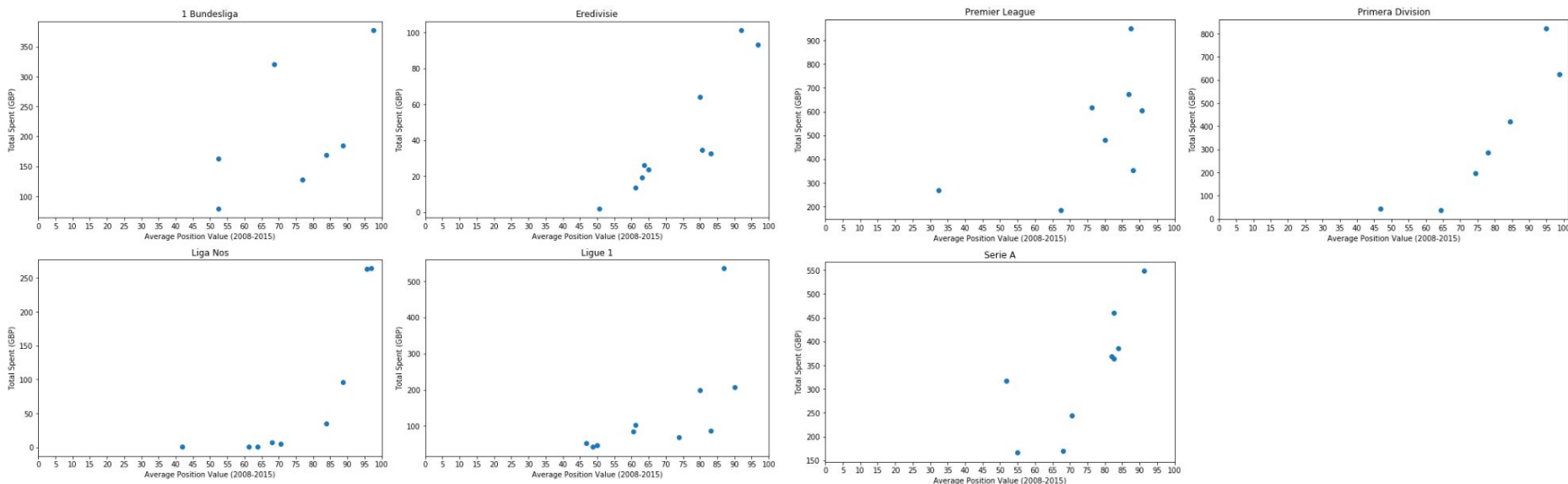
Position Value by Year - Top 5 Teams Per League

- For the top 5 teams in each league, while the top spender doesn't always win the league, spending seems to put a team at least in contention to win in a given year, though there are some exceptions
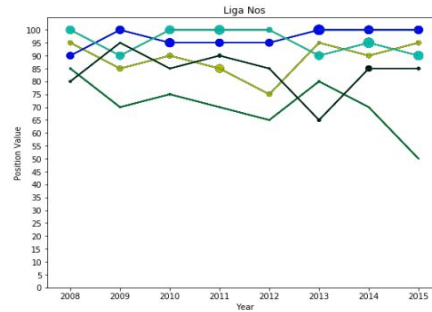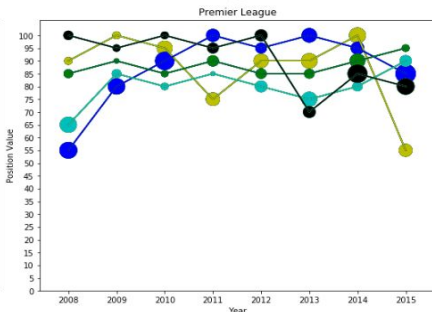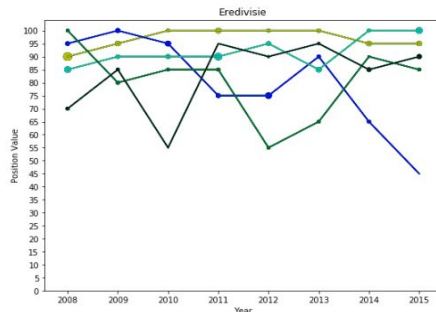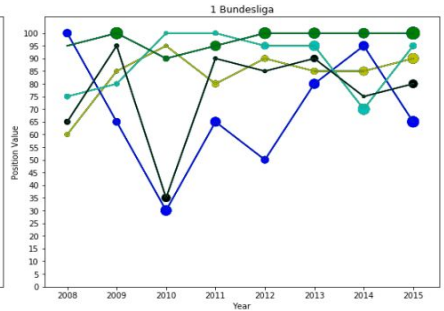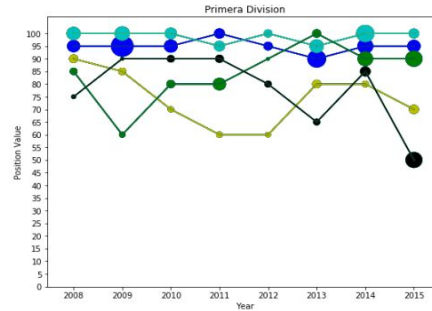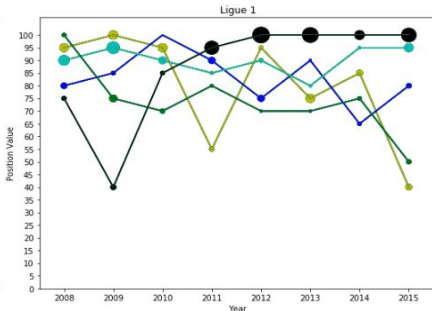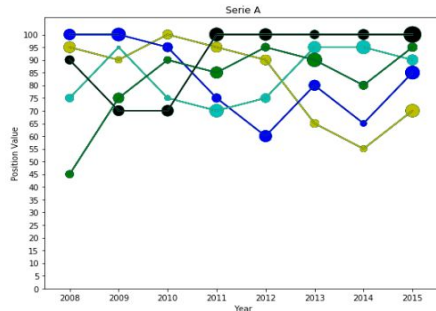
OLS Models

- There was a weak correlation between position value and each spending category (by position and age) in given year
  - Though slightly higher than current, the correlation between position value and spending in previous 2 seasons or 3 seasons, respectively, was still weak
- Much higher correlation between position value and performance categories (win percentage, goals scored/against)
- Despite weak correlation, the p-values of each independent variable were very low, leading to the conclusion there is at least some effect on the dependent variable

In-Depth Analysis

- Applying a Lasso regression shrunk every variable to zero except for win percentage

- Random forest regression showed win percentage the most as variable, but goal difference can also be considered important

Conclusion

- While it looks as though spending money can play a factor for certain clubs, there is no significant statistical relationship overall between where a team finishes in the league and the amount of money they spend.